KATHOLIEKE UNIVERSITEIT
LEUVEN

Faculty of Economics and Applied Economics

Robust M-estimation of multivariate conditionally heteroscedastic time series models with elliptical innovations

Kris Boudt and Christophe Croux

DEPARTMENT OF DECISION SCIENCES AND INFORMATION MANAGEMENT (KBI)

KBI 0727

Robust M-Estimation of Multivariate Conditionally Heteroscedastic Time Series Models with Elliptical Innovations

Kris Boudt* Christophe Croux
Faculty of Economics and Management, K.U.Leuven, Belgium

Abstract

This paper proposes new methods for the econometric analysis of outlier contaminated multivariate conditionally heteroscedastic time series. Robust alternatives to the Gaussian quasi-maximum likelihood estimator are presented. Under elliptical symmetry of the innovation vector, consistency results for M-estimation of the general conditional heteroscedasticity model are obtained. We also propose a robust estimator for the cross-correlation matrix and a diagnostic check for correct specification of the innovation density function. In a Monte Carlo experiment, the effect of outliers on different types of M-estimators is studied. We conclude with a financial application in which these new tools are used to analyse and estimate the symmetric BEKK model for the 1980-2006 series of weekly returns on the Nasdaq and NYSE composite indices. For this dataset, robust estimators are needed to cope with the outlying returns corresponding to the stock market crash in 1987 and the burst of the dotcom bubble in 2000.

Keywords: conditional heteroscedasticity, M-estimators, multivariate time series, outliers, quasi-maximum likelihood, robust methods.

*Correspondence to: Kris Boudt, Faculty of Economics and Management, 69 Naamsestraat, B-3000 Leuven, Belgium. E-mail: Kris.Boudt@econ.kuleuven.be Tel: +32 16 326728 Fax: +32 16 326732

1 Introduction

It is widely recognized that the volatility and the correlations of daily and weekly financial returns may be time-varying. There is no doubt that incorporating these features into the estimation of the conditional covariance matrix of financial returns can lead to better decisions on portfolio optimization, asset pricing and risk management. One way to do this is to specify the conditional covariance matrix as a measurable function of the past of the time series. This approach defines the class of multivariate conditionally heteroscedastic time series models. Most popular is the family of multivariate GARCH (MGARCH) models (see Bauwens *et al.*, 2006, for a recent survey).

It is common to estimate MGARCH models using a Maximum Likelihood (ML) procedure, assuming the innovations to be conditionally Gaussian distributed. Jeantheau (1998) shows that this approach can produce consistent parameter estimates, even when the true distribution is not Gaussian. This is an important result, since it is common to find that after correcting the returns for the dynamics in the conditional covariance matrix, the marginal distribution of the standardized return series is still heavy tailed. An alternative approach is to do ML estimation under the multivariate Student t distribution (Fiorentini *et al.*, 2003). Newey and Steigerwald (1997) prove consistency of the Student t Quasi-Maximum Likelihood (QML) estimator for univariate conditional heteroscedasticity models, provided the true innovation density is unimodal and symmetric around zero.

According to Franses and Ghijsels (1999) the leptokurtosis in the standardized returns can also be interpreted as an indication of outliers in the data set. The effect of outliers on estimating *univariate* GARCH models has been studied in the literature and robust estimators have been proposed. One approach to estimating GARCH models in the presence of outliers is to prune iteratively the outliers and fit the model to the remaining data until no more outliers are detected (Franses and Ghijsels, 1999). Another approach is to use an estimator that is robust to outliers. Sakata and White (1998), Park (2002) and Peng and Yao (2003) estimate the GARCH model by minimizing a robust measure of scale of the residuals. Mancini *et al.* (2005) and Muler and Yohai (2002,2006) propose a robust M-estimator that assigns a much lower weight to outliers than the Gaussian ML estimator does.

All authors find that in the presence of outliers, the Gaussian ML estimator collapses and that robust estimators are more reliable. Van Dijk *et al.* (1999)

find that LM-tests for conditional heteroscedasticity perform better when they are based on robust parameter estimates rather than on the Gaussian QML estimates. Muler and Yohai (2002,2006) argue that in the presence of outliers, it may be more prudent to specify the conditional variance as a bounded function of past returns. Doing so, one limits the impact of outliers on subsequent conditional variance predictions which depend in an autoregressive way on past returns.

To our best knowledge, we are the first to propose robust methods for the econometric analysis of an outlier contaminated *multivariate* conditionally heteroscedastic time series. As compared with their univariate counterparts, multivariate volatility models enable the specification, estimation and evaluation of time-varying correlations and can be used to improve accuracy of forecasts. We use results and techniques from the robustness literature (see e.g. Maronna *et al.*, 2006) to propose robust estimators for the parameter vector underlying the multivariate conditionally heteroscedastic time series model. The new estimators are of the same form as their classical counterpart but give a reduced weight to observations from the extreme tails of the outlier-contaminated distribution.

This paper also contributes to our understanding of the impact of large returns on stock market volatility and on market interdependencies. We robustly estimate the symmetric BEKK volatility model (Engle and Kroner, 1995) for the weekly returns on the Nasdaq and NYSE composite indices between January 1980 and December 2006. We examine the constant correlation hypothesis in the presence of large shared volatility shocks (such as the stock market crash in 1987) and market-specific volatility shocks (such as the burst of the dotcom bubble in 2000).

The remainder of the paper is organized as follows. In Section 2 we prove consistency of M-estimators of conditionally heteroscedastic time series models when the innovations are elliptically distributed. In Section 3 we follow the argument of Muler and Yohai (2002,2006) that QML estimators can be made more robust by bounding their loss function and by bounding the impact of news on the conditional covariance matrix process. Two new diagnostic checks for correct model specification are presented in Section 4. Sections 5-7 illustrate the new methodology through a Monte Carlo study and a financial application. Section 8 summarizes our conclusions and outlines the implications for further research.

2 M-estimation

2.1 Definition

Our analysis starts from a multivariate conditionally heteroscedastic time series model $\mathcal{P}_\theta$ for the centered N-variate random process $\{r_t\}$. Let $I_{r_{t-1}}$ be the information incorporated in the past of r_t. We assume that under $\mathcal{P}_\theta$ the probability of observing r_t conditional on its past is described by the elliptic density function

$$p(r_t|I_{r_{t-1}};\theta) = (\det H_{t,\theta})^{-1/2}\; g\left(r_t'\, H_{t,\theta}^{-1}\, r_t\right), \tag{2.1}$$

where the function $g : \mathbb{R}^+ \to \mathbb{R}^+$. The scatter matrix $H_{t,\theta}$ and the function $g(\cdot)$ have been standardized such that $H_{t,\theta}$ is the conditional covariance matrix of r_t. While there exist elliptic distributions without finite second moment, we consider here only distributions for which the covariance matrix exists. A common choice of $g(\cdot)$ is the standard normal density function which we denote by $\phi(\cdot)$. Another popular choice for $g(\cdot)$ is the standardized Student t density function with $\nu > 2$ degrees of freedom:

$$t_\nu(z) = \frac{\Gamma\left(\frac{\nu+N}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\left[\pi(\nu-2)\right]^{\frac{N}{2}}}\left[1+\frac{z}{\nu-2}\right]^{-\frac{N+\nu}{2}}, \tag{2.2}$$

where $\Gamma(\cdot)$ is the gamma function.

Under $\mathcal{P}_\theta$, the conditional covariance matrix $H_{t,\theta}$ is parameterized by an econometric model as a measurable function of the past of r_t. More precisely, we assume that

$$H_{t,\theta} = H_\theta(r_{t-1}, r_{t-2}, ...), \tag{2.3}$$

with $H_\theta(\cdot)$ a $N \times N$ positive definite, symmetric matrix function of a sequence of N-dimensional vectors. As an illustration, consider the BEKK parameterization for $H_{t,\theta}$ (Engle and Kroner, 1995):

$$H_{t,\theta} = C'C + A'r_{t-1}r_{t-1}'A + B'\,H_{t-1,\theta}\,B, \tag{2.4}$$

where A, B and C are $N \times N$ parameter matrices and C is upper triangular.

Denote θ_* the true, unknown parameter vector belonging to the parameter space Θ and $g_*(\cdot)$ the true density function in (2.1). It is common to estimate θ_* by the value of $\theta \in \Theta$ that maximizes the joint density of the sample

$S_T = \{r_1, ..., r_T\}$ constructed under the nominal assumption that the density function of the data is $g(\cdot)$. This estimation approach yields the QML estimator

$$\hat{\theta}_{QML} \;=\; \text{argmax}_{\theta\in\Theta} \; - \underset{t}{\text{ave}} \left[\log\det H_{t,\theta} \;-\; 2\log g\left(r_t' \, H_{t,\theta}^{-1} \, r_t\right)\right], \tag{2.5}$$

where ave_t stands for the arithmetic average over $t = 1, ..., T$. It coincides with the ML estimator when $g(\cdot) = g_*(\cdot)$. Jeantheau (1998) proves consistency of the Gaussian QML estimator, which is given by

$$\hat{\theta}_{\phi-QML} \;=\; \text{argmax}_{\theta\in\Theta} \; - \underset{t}{\text{ave}} \left[\log\det H_{t,\theta} \;+\; r_t' \, H_{t,\theta}^{-1} \, r_t\right]. \tag{2.6}$$

The QML estimator belongs to the broader class of M-estimators defined as follows.

Definition 1 *The M-estimate based on a sample $S_T = \{r_1, ..., r_T\}$ is the value of $\theta \in \Theta$ for which the M-function*

$$M\left(S_T; \theta, \rho\right) = \underset{t}{\text{ave}} \;\left[\log\det H_{t,\theta} \;+\; \sigma_* \rho\left(r_t' \, H_{t,\theta}^{-1} \, r_t\right)\right], \tag{2.7}$$

is minimized, with $\rho(\cdot)$ a positive, non-decreasing scalar function. The scalar σ_* is a consistency factor.

The consistency factor σ_* depends on the true density function $g_*(\cdot)$ and on $\psi(\cdot) = \rho'(\cdot)$, the first derivative of the function $\rho(\cdot)$. It equals

$$\sigma_* = N \left\{ \frac{2\pi^{N/2}}{\Gamma(N/2)} \int_0^\infty \psi(t^2) t^{N+1} g_*(t^2) dt \right\}^{-1}. \tag{2.8}$$

In Subsection 3.1 we present numerical values for σ_*.

The function $\rho(\cdot)$ in the above definition is called the loss function associated to the M-estimator. If $\rho(\cdot) = -2\log g(\cdot)$, with $g(\cdot)$ a specified density function, we obtain a QML estimator. The Gaussian ϕ and standardized Student t_ν density functions are the most popular elliptical density functions used to describe financial return series. The corresponding Gaussian and Student t_ν $(\nu > 2)$ loss functions are given by

$$\rho_\phi(z) = z \quad ; \quad \rho_{t_\nu}(z) = (N+\nu)\log\left[1 + \frac{z}{\nu - 2}\right]. \tag{2.9}$$

These loss functions are unbounded. In Section 3 we will discuss loss functions that are bounded and do not correspond to a specified distribution of the innovation vectors.

2.2 Asymptotic properties

Jeantheau (1998)'s proof of consistency of the Gaussian QML estimator can only be used to prove consistency of M-estimators whose loss function $\rho(\cdot)$ satisfies $E_{\theta_*}\rho(z) = \rho(E_{\theta_*}z)$, where E_{θ_*} denotes the expectation under the true model $\mathcal{P}_{\theta_*}$. Moreover it requires that the covariance matrix of the innovation vector exists. In Appendix A we prove consistency for a wide range of M-estimators, including M-estimators with a Student t_ν loss function. Denote $m(r_t;\theta,\rho) = \log\det H_{t,\theta} + \sigma_*\rho\left(r_t' H_{t,\theta}^{-1} r_t\right)$. The consistency result is obtained under the following assumptions on the loss function, on the derivative of the M-function $\dot{M}(r_t;\theta,\rho) = \partial M(r_t;\theta,\rho)/\partial\theta$ and on the density function of the innovations.

A1 *The loss function $\rho(\cdot)$ has the following properties.*

1. *Its derivative $\psi(\cdot) = \rho'(\cdot)$ is nonnegative, nonincreasing and continuous.*
2. *The function $\psi(z)z$ is bounded. Let $K = \sup_{z\geq 0}\psi(z)z$.*
3. *The function $\psi(z)z$ is nondecreasing, and is strictly increasing in the interval where $\psi(z)z < K$.*
4. *There exists z_0 such that $\psi(z_0^2)z_0^2 > N$ and that $\psi(z)z > 0$ for $z \leq z_0$.*
5. *There exists $a > 0$ such that for every hyperplane H, the probability under $\mathcal{P}_{\theta_*}$ for an observation to lie in H is at most $1 - N/K - a$.*

A2 $p(r_t|I_{r_{t-1}};\theta_*) = (\det H_{t,\theta_*})^{-1/2}\; g_*\left(r_t' H_{t,\theta_*}^{-1} r_t\right)$.

A3 $H_{r_t,\theta} = H_{r_t,\theta_*} \Rightarrow \theta = \theta_*$.

A4 Θ *is compact.*

A5 $E_{\theta_*}[\partial m(r_t;\theta,\rho)/\partial\theta] < \infty$.

A6 $\operatorname{plim}\dot{M}(S_T;\theta,\rho) = 0 \Rightarrow E_{\theta_*}[\sigma_*\psi(r_t' H_{t,\theta}^{-1} r_t) r_t r_t' | I_{r_{t-1}}] = H_{t,\theta}$.

Proposition 1 *Under A1-A6 the M-estimator $\hat{\theta}$ defined in Definition 1 is a consistent estimator for θ_*.*

An intuitive interpretation of this result is that under the previous assumptions, minimizing the M-function coincides with finding the value of $\theta \in \Theta$ for which the estimated conditional covariance matrix equals the conditional expectation of a weighted covariance matrix,

$$H_{t,\theta} = E_{\theta_*}\left[\sigma_* \psi(r_t' \, H_{t,\theta}^{-1} \, r_t) r_t r_t' | I_{r_{t-1}}\right]. \tag{2.10}$$

The weights are proportional to $\psi(r_t' \, H_{t,\theta}^{-1} \, r_t)$, with $d_{t,\theta}^2 = r_t' \, H_{t,\theta}^{-1} \, r_t$ the squared Mahalanobis distance of r_t. We have that asymptotically, under A1-A6, θ_* is the unique solution to this problem[1].

Under regularity conditions, M-estimators are asymptotically normal with asymptotic covariance matrix (Chapter 7 in Hayashi, 2000):

$$\left\{ E_{\theta_*}\left[\frac{\partial^2 m(r_t; \theta_*, \rho)}{\partial \theta \partial \theta'}\right] \right\}^{-1} \Sigma \, \left\{ E_{\theta_*}\left[\frac{\partial^2 m(r_t; \theta_*, \rho)}{\partial \theta \partial \theta'}\right] \right\}^{-1}, \tag{2.11}$$

where Σ is the long run covariance matrix of $\partial m(r_t; \theta_*, \rho)/\partial \theta$. In the empirical application of Section 7, we estimate this quantity by its sample counterpart whereby the Bartlett HAC long-run autocovariance matrix estimate is used as well as the formulas in Hafner and Herwartz (2003) for computing the score and Hessian of the M-function analytically.

3 Robust M-estimation

For many applications, a fully specified conditionally heteroscedastic time series model with elliptical innovations is at most a very good approximation of the true data generating process. It may explain well the bulk of the observations, but real-world time series will almost always contain some outlying observations or discrepant substructures. In a conditionally heteroscedastic time series setting, outliers can be defined as observations that are extremely unusual given the past of the series. It is common to do as if there were no outliers in the data and to estimate θ_* by the value of $\theta \in \Theta$ that minimizes the M-function (2.7). Such an approach will still yield reliable result in the presence of outliers if (1) a bounded loss function or a loss function with a derivative that decreases sufficiently fast towards zero, is used; (2) a specification for $H_{t,\theta}$ is taken under which innovations have a limited impact on the conditional covariance matrix process.

[1] Note that this result does not require that the distribution of the innovations has a finite second moment.

3.1 The loss function

Outliers are characterized by a large value of $d^2_{t,\theta_*} = r_t' \, H^{-1}_{t,\theta_*} \, r_t$, which is the squared Mahalanobis distance of r_t under the true model. The shape of the loss function associated to the M-estimator is very important since it controls the impact of $d^2_{t,\theta}$ on the M-function in (2.7). Observe in (2.9) that outliers have an unbounded impact on the objective function of the M-estimator with Gaussian or Student t_ν loss function because these loss functions are unbounded. Hence, a bounded loss function is called for.

From (2.10) it follows also that the shape of $\psi(\cdot)$, the first derivative of the loss function $\rho(\cdot)$, is important. Indeed, it determines the weight given to each observation as a function of $d^2_{t,\theta}$. Robust M-estimators will be characterized by a loss function whose derivative, for large values of $d^2_{t,\theta}$, decreases sufficiently fast and tends to zero at infinity. The derivatives of the Gaussian and Student t_ν loss functions equal

$$\rho'_\phi(z) = 1 \quad ; \quad \rho'_{t_\nu}(z) = \frac{N+\nu}{\nu - 2 + z}. \tag{3.1}$$

We see that the M-estimator with Gaussian loss function is not robust because it gives the same weight to all realizations irrespective of their degree of outlyingness, as measured by $d^2_{t,\theta}$. However, the derivative of the Student t_ν loss functon is decreasing and tends to zero at infinity. The smaller ν is, the more robust the M-estimator with Student t_ν loss function will be.

A proposal of a loss function that combines the desirable properties of boundedness and decreasing first derivative equals the Gaussian or Student t_ν loss function for reasonable values of r_t and is bounded for extreme values of r_t. This approach is pursued by Muler and Yohai (2002,2006) for univariate GARCH processes. An elegant way of creating such a loss function is to make use of an operator, say ω, that projects any function $\rho(\cdot)$ on its bounded version $\omega_\rho(\cdot)$. We define the operator ω as follows:

$$\omega_\rho(z) = \begin{cases} \rho(z) & \text{for} \quad 0 \leq z < \chi^2_{N,\alpha_1} \\ q_\rho(z) & \text{for} \quad \chi^2_{N,\alpha_1} \leq z < \chi^2_{N,\alpha_2} \\ q_\rho(\chi^2_{N,\alpha_2}) & \text{for} \quad z \geq \chi^2_{N,\alpha_2}. \end{cases} \tag{3.2}$$

Here $z \geq 0$ and $q_\rho(\cdot)$ is the unique quadratic function for which $\omega_\rho(\cdot)$ and $\omega'_\rho(\cdot)$ are continuous. The threshold $\chi^2_{N,\alpha}$ is the α-quantile of the chi-squared distribution with N degrees of freedom, which is the distribution function of d^2_{t,θ_*} when r_t is conditionally normal. For most density functions, we have

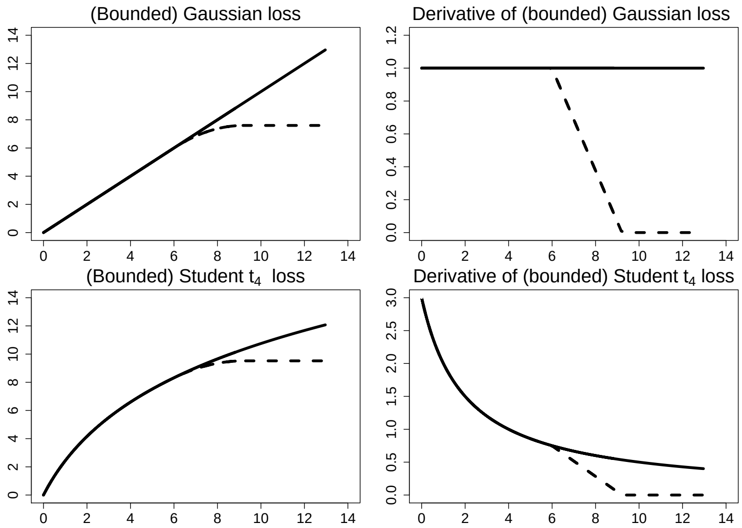

Figure 1: Gaussian and Student t_4 loss functions and their derivative (solid line), together with their bounded counterparts (dashed line).

that the higher α_1, α_2 are, the more efficient the estimate is in the absence of outliers and the lower they are, the more robust it is. Also α_1 may not be too close to α_2; otherwise the slope of $\omega_\rho(\cdot)$ between χ^2_{N,α_1} and χ^2_{N,α_2} will be too steep, which has a harmful effect on the M-estimator. We set $\alpha_1 = 0.95$ and $\alpha_2 = 0.99$. The Gaussian and Student t_4 loss functions and their derivative are plotted in Figure 1, together with their bounded counterpart and its derivative. Note that, by construction, the derivative of the bounded loss function gives a zero weight to extreme observations.

Table 1 reports the consistency factors for the M-estimators with (un)bounded Gaussian and Student t_4 loss functions, obtained by numerical integration of the expression (2.8). Observe that for the M-estimator with Gaussian loss function (being the Gaussian QML estimator) the consistency factor is one whatever be the true density function (g_*) and the dimension of the data (N). This is not true for the M-estimator with Student t_ν loss function whose consistency factor increases with the dimension of the series and the thickness of the tails of the true density function. The consistency factor is one when the loss function is proportional to $\log g_*$. In this case, the M-estimator coincides with the ML estimator. The consistency factor of the M-estimator with bounded loss function is always larger than the one of its unbounded counterpart.

g_*	ϕ			t_6			t_4		
N	1	2	3	1	2	3	1	2	3
ρ_ϕ	1	1	1	1	1	1	1	1	1
ω_{ρ_ϕ}	1.129	1.104	1.090	1.386	1.384	1.392	1.603	1.612	1.630
ρ_{t_4}	0.826	0.831	0.838	0.915	0.918	0.922	1	1	1
$\omega_{\rho_{t_4}}$	0.865	0.863	0.867	1.009	1.169	1.335	1.124	1.297	1.475

Table 1: Consistency factors for M-estimators with (un)bounded Gaussian and Student t_4 loss function. The consistency factor is computed for the density function g_* of a series of dimension N.

3.2 Impact of news on subsequent volatility predictions

Most conditionally heteroscedastic time series models specify the elements of the conditional covariance matrix as an unbounded function of a distributed lag of the squares and cross-products of the components of the innovation vector r_t. It follows that the news impact curve, which measures the effect of the innovations on the one-step ahead conditional covariance matrix (Engle and Ng, 1993), is unbounded. Moreover, because of the autoregressive structure of the process, the impact of news propagates to future values of the conditional covariance matrix process. Muler and Yohai (2002,2006) stress that because of this property of innovation propagation, neglecting outliers renders any inference on autoregressive time series processes extremely sensitive to outliers. For GARCH processes this is aggravated by the fact that the news impact curve is typically quadratic. Hence, neglected outliers may have a very adverse effect on subsequent values of the estimated volatility process.

It is possible to bound the impact of outliers on the conditional covariance matrix by specifying it as a function $H_\theta(\cdot)$ of the *weighted* observations $\tilde{r}_t = w(d^2_{t,\theta})r_t$ instead of the raw observations r_t. The weight function $w(\cdot)$ downweights extreme observations, that is observations with a large value of $d^2_{t,\theta}$. When $H_\theta(\cdot)$ has the BEKK specification (2.4), we obtain the following specification for the conditional covariance matrix:

$$\tilde{H}_{t,\theta} = H_\theta(\tilde{r}_{t-1}, \tilde{r}_{t-2}, ...) = C'C + w^2(d^2_{t-1,\theta})A'r_{t-1}r'_{t-1}A + B'\,\tilde{H}_{t-1,\theta}\,B. \quad (3.3)$$

The weight function we will use in the application is

$$w(d^2_{t,\theta}) = [\omega_f(d^2_{t,\theta})/d^2_{t,\theta}]^{\frac{1}{2}}, \quad (3.4)$$

where $f(z) = z$ and $\omega_f(\cdot)$ is as defined in (3.2), with $\alpha_1 = 0.95$ and $\alpha_2 = 0.99$. This weight function is plotted in Figure 2 for $N = 2$. We see that only the observations with squared distance above the 95% quantile of the χ^2_2 distribution are downweighted. The weight function tends to zero at infinity such that the impact of outliers on the conditional covariance matrix is limited. The idea of robustifying the conditional covariance matrix against outliers in the data by downweighting past observations of r_t in the GARCH specification has already been pursued by Muler and Yohai (2002,2006) in the univariate case. We follow them in calling this the *Bounded Innovation Propagation* (BIP) conditionally heteroscedastic time series model.

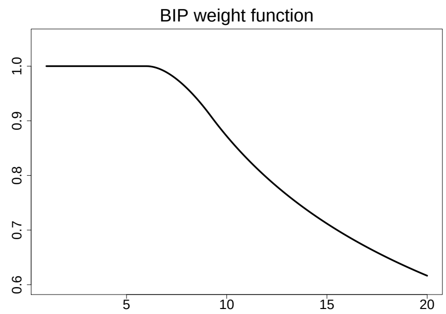

Figure 2: Weight function used in the Bounded Innovation Propagation (BIP) conditionally heteroscedastic time series model.

Muler and Yohai (2002,2006) estimate the parameters of the univariate GARCH model by minimizing the M-function corresponding to the BIP-GARCH model and call the newly obtained estimator the BIP M-estimator. This approach readily extends to the multivariate case in which the BIP M-estimator is defined as the value of $\theta \in \Theta$ that minimizes the BIP M-function

$$\tilde{M}\left(S_T; \theta, \rho\right) = \underset{t}{\text{ave}} \left[\log\det \tilde{H}_{t,\theta} + \sigma_* \rho\left(r_t' \, \tilde{H}_{t,\theta}^{-1} \, r_t\right)\right]. \tag{3.5}$$

Because the conditional covariance matrix is misspecified under $\mathcal{P}_{\theta_*}$, the BIP M-estimator may not be consistent for θ_* in the absence of outliers. In practice, the data will not follow $\mathcal{P}_{\theta_*}$ exactly and the asymptotic bias of the BIP M-estimator under $\mathcal{P}_{\theta_*}$ may be compensated by its reduced sensitivity to outliers in the data.

4 Diagnostic checks

Consistency of M-estimators has been proven under the assumption of correct specification of the conditional mean and covariance matrix of the time series. The robustly estimated cross-correlation in the series is a helpful tool for appropriate specification of the conditional moments of the data. In Subsection 4.1 we propose a robust estimator for this quantity. Another issue is that the definition of a consistent M-estimator depends on knowledge of the true density function $g_*(\cdot)$ through the consistency factor σ_* in (2.8). In Subsection 4.2 a diagnostic check for correct specification of $g(\cdot)$ is proposed.

4.1 A robust sample cross-correlation matrix

Prior to estimation of a conditionally heteroscedastic time series model, it is useful to check whether the series present evidence of conditional heteroscedasticity. To our knowledge, there are no robust multivariate checks for conditional heteroscedasticity that do not require prior estimation of the model. In the empirical application, we will use the following multivariate extension of the α-trimmed autocorrelation proposed by Chan and Wei (1992).

Definition 2 *Let $a_1, ..., a_T$ be a N-dimensional vector time series of length T and $0 < \alpha < 1$ a trimming factor. The α-trimmed lag l sample cross-correlation matrix is the $N \times N$ matrix which (i, j)-th element is given by*

$$\frac{\sum_{t=l+1}^{T}(a_{i,t} - \overline{a}_i)(a_{j,t-l} - \overline{a}_j)L_t^\alpha L_{t-l}^\alpha}{\left[\sum_{t=l+1}^{T}(a_{i,t} - \overline{a}_i)^2 L_{t-l}^\alpha L_t^\alpha \sum_{t=l+1}^{T}(a_{j,t} - \overline{a}_j)^2 L_{t-l}^\alpha L_t^\alpha\right]^{1/2}}, \tag{4.1}$$

with indicator function

$$L_t^\alpha = \begin{cases} 0 & \text{if } (a_t - \overline{a})'\Sigma_t^{-1}(a_t - \overline{a}) \leq \chi^2_{N,1-\alpha}, \\ 1 & \text{otherwise.} \end{cases} \tag{4.2}$$

The N-dimensional vector $\overline{a}$ and square matrix Σ_t are robust measures of the mean and local covariance matrix of a_t, respectively.

Throughout this paper we will use Rousseeuw (1985)'s 75% minimum covariance determinant (MCD) estimator for estimating the unconditional mean and covariance matrix of multivariate time series. They are defined as the sample moments of the subset for which the covariance matrix has the lowest determinant among all subsets containing 75% of the observations.

The implicit outlier detection criterion used to compute the α-trimmed lag l sample cross-correlation matrix differs from the one usually used in multivariate outlier detection methods (see e.g. Rousseeuw and van Zomeren, 1990) by measuring distances with respect to a robust measure of the *local* covariance matrix of the series. Doing so, it avoids the masking of outliers in periods of low volatility and the identification of regular observations as outliers (called swamping) in periods of high volatility. The conditional mean of the series is assumed to be constant over time. In the application we set $\alpha = 0.01$ and take Σ_t as the MCD covariance estimate applied to the observations in the time window $t-2, ..., t+2$.

Because the lag l cross-correlation is independent of the trimming criterion, the distribution theory for the classical sample cross-correlation (Tiao and Box, 1981) carries over to this estimator. A useful result is that when a_t is white noise, the cross-correlation of a_t is asymptotically normal with mean zero and variance $1/\sum_{t=k+1}^{T} L_{t-k}^{\alpha} L_t^{\alpha}$, the inverse of the number of observations used to compute the trimmed autocorrelation estimate. Note that the effect of trimming is a decrease in the efficiency of the estimator and an increase in its robustness. In the financial application in Section 7 we use this result to detect any significant cross-correlations in the return and absolute return series of two major stock indices.

4.2 Consistency factor test statistic

An issue in applying the M-estimator in practice is that it requires knowledge of the true density function through the definition of the consistency factor σ_*. Here we propose a test for the null that $g(\cdot) = g_*(\cdot)$. Under the null, we have that the expression in (2.10) will only hold for θ_*. We can rewrite (2.10) as follows:

$$E_{\theta_*}\left[\sigma_* \psi(r_t' H_{t,\theta}^{-1} r_t) r_t r_t' H_{t,\theta}^{-1} | I_{r_{t-1}}\right] = I_N. \tag{4.3}$$

Taking the trace, we find that if $\theta = \theta_*$, then the following result must hold:

$$E_{\theta_*}\left[\sigma_* \psi(r_t' H_{t,\theta}^{-1} r_t) r_t' H_{t,\theta}^{-1} r_t | I_{r_{t-1}}\right] = N. \tag{4.4}$$

Consequently, the following statistic can be used to test for the null of a correctly specified density function.

Definition 3 *The consistency factor test statistic for testing $H_0 : g(\cdot) = g_*(\cdot)$ is given by*

$$\kappa(g) = \sqrt{T}\, \frac{\mathrm{ave}_t\, \psi(r_t'\, H_{t,\theta}^{-1}\, r_t) r_t'\, H_{t,\theta}^{-1}\, r_t - m}{s}, \tag{4.5}$$

with $m = \frac{2\pi^{N/2}}{\Gamma(N/2)} \int_0^\infty \psi(t^2) t^{N+1} g(t^2) dt,\ s = \left[\frac{2\pi^{N/2}}{\Gamma(N/2)} \int_0^\infty \left[\psi(t^2)t^2 - m\right]^2 t^{N-1} g(t^2) dt\right]^{\frac{1}{2}}$ *and $\psi(\cdot)$ the derivative of the loss function of the M-estimator used to estimate the unknown θ.*

Since under the model assumptions made previously, the distances are conditionally independent and the M-estimator is consistent, one can invoke a central limit theorem to state that this statistic is asymptotically standard normal distributed when $g(\cdot) = g_*(\cdot)$.

5 Implementation

The aim of this section is twofold. First we specify the particular model within the family of multivariate conditional heteroscedasticity models that will be used in the Monte Carlo study of Section 6 and the financial application in Section 7. Second, we discuss the numerical procedure used to obtain the parameter estimates.

5.1 The model

We consider the bivariate conditionally Student t_ν BEKK volatility model of order one proposed by Engle and Kroner (1995). Under this model, the data generating process is given by the following set of equations:

$$\begin{cases} r_t | I_{r_{t-1}} & \overset{i.i.d.}{\sim} & t_\nu(0, H_{t,\theta})\ ;\ \nu > 2 \\ H_{t,\theta} & = & C'C + A' r_{t-1} r_{t-1}' A + B'\, H_{t-1,\theta}\, B. \end{cases} \tag{5.1}$$

The parameter matrices C, A and B all denote 2×2 coefficient matrices and C is upper triangular. To reduce the number of parameters, we further assume that the matrices A and B are symmetric. Let θ be the vector of length 9 that stacks the upper diagonal elements of the matrices C, A and B one on top of the other. Observe that each element of the conditional covariance matrix is specified as a linear function of the lagged squares and cross-products of the elements of r_t as well as of the lags of the elements of the conditional covariance matrix. As such, the BEKK model can replicate the stylized fact that volatility clusters over time and co-moves across assets.

The way in which the conditional covariance matrix process evolves in time depends on the parameter values and the realizations of the stochastic process $\{r_t\}$. In general, one wants the process $\{H_{t,\theta}\}$ to be positive definite and covariance stationary whatever be the realizations of $\{r_t\}$. The BEKK model thanks its popularity to the fact that this can be guaranteed under simple constraints on the parameter space. Indeed, Engle and Kroner (1995) show that if C is of full rank, $\{H_{t,\theta}\}$ will be positive definite for all t and covariance stationary if and only if all the eigenvalues of $A' \otimes A' + B' \otimes B'$ are less than one in modulus. Because of the quadratic forms, different values of θ may generate the same sequence $\{H_{t,\theta}\}$. In the bivariate case, the following conditions ensure $\{H_{t,\theta}\}$ to be identified, positive definite and covariance stationary:

$$C_{11} > 0, C_{22} > 0, A_{11} > 0, B_{11} > 0 \tag{5.2}$$

$$\zeta_\theta \equiv \zeta\left(A' \otimes A' \ + \ B' \otimes B'\right) < 1, \tag{5.3}$$

where $\zeta(\cdot)$ is the operator that takes the largest eigenvalue in absolute value of its argument. In the following, it is assumed that the true parameter vector θ_* belongs to the parameter space Θ defined as the subset of $\mathbb{R}^9$ for which the conditions (5.2)-(5.3) are satisfied.

5.2 Algorithm

The M-estimators for the bivariate BEKK volatility model are defined as the solution to a highly nonlinear minimization problem, without an explicit solution. The algorithm we use to compute the M-estimates proceeds in three steps.

First the raw time series is centered by subtracting the robust minimum covariance determinant (MCD) estimate of location. The recursive conditional covariance matrix process is initialized at the MCD estimate for the unconditional covariance matrix. Second, starting values for the parameter estimates are found. In the Monte Carlo experiment, we take the true parameter vector as starting value for the parameter estimates. For the empirical application, a grid search is used to obtain a reasonable set of starting values. Finally we switch to a quasi-Newton minimization algorithm in which the score and the Hessian are computed on the basis of finite difference procedures. We impose the constraints (5.2) and (5.3) through a special type of active set method, called the gradient projection method (see e.g. Nocedal and Wright, 2000). Because this method cannot deal with the very complex and non-linear covariance stationarity condition (5.3), we only impose the following necessary conditions for (5.3) to be satisfied:

$$\begin{gathered} |A_{11}| < 1, |A_{22}| < 1, |B_{11}| < 1, |B_{22}| < 1 \\ A_{11}^2 + B_{11}^2 < 1, A_{11}A_{22} + B_{11}B_{22} < 1, A_{22}^2 + B_{22}^2 < 1. \end{gathered} \tag{5.4}$$

These conditions can be derived by noting that for the bivariate symmetric BEKK model the covariance stationarity constraint in (5.3) is equivalent to the condition that $I_4 - A' \otimes A' - B' \otimes B'$ is a positive definite matrix (Altay-Salih *et al.*, 2003).

6 Monte Carlo

This section aims to compare the behavior of the (BIP) M-estimators with bounded and unbounded Gaussian and Student t_4 loss function for simulated date with different levels of outlier contamination. The artificial time series are generated as follows. First outlier-free time series $\{y_t\}$ of length 1000 are generated from model (5.1), with parameter matrices:

$$C'C = \begin{pmatrix} 0.16 & 0.20 \\ 0.20 & 0.34 \end{pmatrix} \; A = \begin{pmatrix} 0.25 & 0.04 \\ 0.04 & 0.30 \end{pmatrix} \; B = \begin{pmatrix} 0.85 & 0.05 \\ 0.05 & 0.80 \end{pmatrix}.$$

We then add outliers to the clean series using the following probabilistic setup:

$$r_t = y_t + 5\xi_t d_t. \tag{6.1}$$

Under this model the occurrence of an outlier in any of the two series is governed by the random process $\{d_t\}$, modeled as a sequence of i.i.d. draws from a Bernoulli distribution where the occurrence of an outlier ($d_t = 1$) has probability ε. The transmission and the size of the outliers stems from a 2-dimensional i.i.d. vector process $\{\xi_t\}$ constructed such that, when an outlier occurs, it will be either in the first component, second component or both with probability 0.3, 0.3 and 0.4, respectively. The magnitude of the outlier is 5 times the conditional standard deviation of the corresponding element of y_t.

Since M-estimators are defined by the extremum of the M-function, it is important that the location of the extremum as a function of the parameter vector is robust to outliers in the data. In Figure 3 the M-function (2.7) with (bounded) Gaussian and Student t_4 loss function are plotted as a function of θ_9, once in the absence and once in the presence of 5% of outliers. The outlier-free time series is conditionally Gaussian distributed. We see that in the absence of outliers, all M-functions reach an extremum close to the true parameter value, which is 0.8. Because of the quadratic forms in the BEKK specification, the M-function reaches also a local minimum around $\theta_9 = -0.8$. Hence, the importance of having good starting values and of imposing the bound constraints (5.2) in the optimization process.

Note that contaminating 5% of the data with additive outliers increases the variability of the M-Function. The magnitude of this outlier induced variability depends on the loss function. For the Gaussian loss function (ρ_ϕ), the variability of the M-function is so large that there is no more visible evidence that the M-function reaches an extremum close to the true parameter value. Moreover the outliers modify the global shape of the M-function with Gaussian loss function. This is not the case for the two other loss functions. The outlier induced variability is much smaller for the M-function with bounded Gaussian (ω_{ρ_ϕ}) and Student t_4 (ρ_{t_4}) loss function. The global shape of these M-functions is little influenced by the outliers.

Of course this is only a partial analysis. To study the effect of outlier contamination on the bias and efficiency of the (BIP) M-estimator with (un)bounded Gaussian or Student t_4 loss function (a total of 8 estimation procedures), we consider 400 replications of (6.1), three levels of outlier contamination ($\varepsilon = 0$, $\varepsilon = 0.01$ and $\varepsilon = 0.05$) and innovations that are either conditionally Gaussian or Student t_4 distributed. For each M-estimator and for each type of simulated series, we aggregate the parameter estimates to compute the mean error (ME) and the root mean squared error (RMSE) as follows,

$$\text{ME} = \| \underset{j}{\text{ave}}\, \hat{\theta}_j - \theta_* \| \ , \ \text{RMSE} = \sqrt{\underset{j}{\text{ave}}\, \|\hat{\theta}_j - \theta_*\|^2}, \tag{6.2}$$

where ave_j denotes the average across the 400 replications and $\| \cdot \|$ is the Euclidian norm operator. These summary statistics are reported in Table 2.

We find that in the absence of outliers ($\varepsilon = 0$), the estimation of the BEKK model on the basis of the BIP-BEKK model does not seem to have a large impact on the bias of the M-estimator. Under Gaussian innovations, the finite sample bias and efficiency of the M -and BIP M-estimator with the same loss function are similar and under Student t_4 innovations, the M-estimator

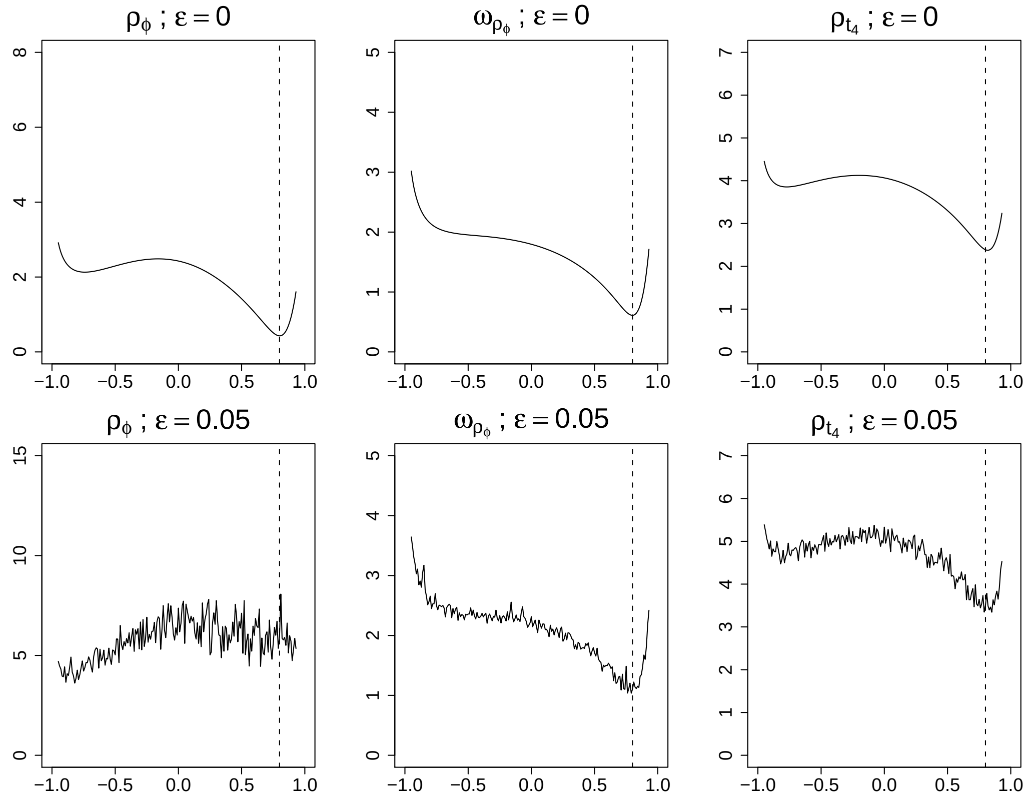

Figure 3: M-functions with unbounded Gaussian (ρ_ϕ), bounded Gaussian (ω_{ρ_ϕ}) and unbounded Student t_4 (ρ_{t_4}) loss function for different levels of contamination ε. Except for θ_9, the M-function is computed at the true parameter values, which is 0.8 for θ_9.

does only slightly better[2].

When the series are outlier contaminated ($\varepsilon = 0.01$ or $\varepsilon = 0.05$), an overall result is that outliers adversely affect the bias as well as the efficiency of all M-estimators. Consistent with Sakata and White (1998), we find that ML estimation under a heavy-tailed density function does not yield outlier robust estimates. Indeed, not only the Gaussian but also the Student t_4 ML estimators collapse in the presence of small levels of outlier contamination. The lack of robustness of the M-estimator with a Student t_4 loss function is because of the unbounded innovation propagation in the conditional covariance matrix process. In the presence of outliers in the data, the BIP version of this

[2]Surprisingly, for Gaussian innovations, the ML estimator does not have the smallest RMSE at finite samples.

estimator needs to be used.

Note that under Gaussian innovations, the (BIP) M-estimator with bounded Student t_4 loss function ($\omega_{\rho_{t_4}}$) does not perform well when 5% of the observations are outlier contaminated. It seems that, provided the impact of news on volatility is bounded, a loss function with a smooth, sufficiently fast decreasing derivative such as $\rho_{t_4}(\cdot)$, is a better characteristic of a robust M-estimator than a bounded loss function. A conclusion of this simulation experiment is that, among the considered estimators, no M-estimator does better than the BIP M-estimator with Student t_4 loss function.

Loss function		ρ_ϕ		ω_{ρ_ϕ}		ρ_{t_4}		$\omega_{\rho_{t_4}}$	
M / BIPM		M	BIPM	M	BIPM	M	BIPM	M	BIPM
$g_* = \phi$									
$\varepsilon = 0$	ME	0.080	0.074	0.092	0.102	0.033	0.032	0.075	0.071
	RMSE	0.188	0.180	0.227	0.242	0.092	0.085	0.202	0.206
$\varepsilon = 0.01$	ME	0.233	0.170	0.198	0.083	0.229	0.036	0.192	0.098
	RMSE	0.482	0.432	0.312	0.206	0.334	0.079	0.332	0.286
$\varepsilon = 0.05$	ME	0.441	0.367	0.279	0.114	0.434	0.064	0.358	0.458
	RMSE	0.629	0.592	0.353	0.305	0.476	0.102	0.574	0.779
$g_* = t_4$									
$\varepsilon = 0$	ME	0.099	0.106	0.148	0.158	0.067	0.075	0.121	0.131
	RMSE	0.245	0.257	0.266	0.280	0.095	0.116	0.225	0.230
$\varepsilon = 0.01$	ME	0.235	0.178	0.220	0.142	0.177	0.072	0.195	0.110
	RMSE	0.488	0.443	0.359	0.267	0.266	0.096	0.314	0.212
$\varepsilon = 0.05$	ME	0.433	0.358	0.299	0.105	0.436	0.108	0.278	0.080
	RMSE	0.644	0.576	0.387	0.218	0.486	0.133	0.337	0.187

Table 2: Mean estimation error (ME) and root mean squared error (RMSE), over 400 replications, for the (BIP) M-estimators with (un)bounded Gaussian and Student t_4 loss function. The simulated series is either conditionally Gaussian ($g_* = \phi$) or Student t_4 ($g_* = t_4$) distributed. The level of contamination is ε.

7 A financial application

There are many applications in finance that rely on an estimate of the conditional covariance matrix of returns. Examples include portfolio allocation, financial risk management and asset pricing decisions. It is well known that due to market crashes and rallies, many financial data sets contain atypical observations. We may thus expect that the robust methods discussed in Sections 3 and 4 can be used to improve decision making. Moreover, because they are simple modifications of the classical ML estimators, they are easy to understand by practicioners. In this section we apply the robust methods to the time series of weekly returns on the Nasdaq and NYSE composite indices.

7.1 Data

We consider the data set of weekly bivariate return vector observations for the Monday close prices of the dollar-denominated Nasdaq and NYSE composite indices. The data source is Datastream. The sample period ranges from January 1980 through December 2006 (1404 observations). All returns are continuously compounded returns and expressed in percentage points. Denote r_t the return vector whose first and second component is the t-th weekly return on the Nasdaq and NYSE composite index, respectively.

The series are plotted in Figure 4. Note the large negative return corresponding to the stock market crash of October 19, 1987. Interestingly, the Nasdaq index had returns of similar magnitude in the year 2000. In an unconditional framework, these returns are qualified as outliers, whereas under a conditionally heteroscedastic time series model, some of these extreme returns can be explained by the model. Observe that not only there is volatility clustering in both series, but also that there is a lot of commonality in volatility across the two series.

Let us first concentrate on the unconditional properties of the time series. Figure 5 reports a scatter plot of the data and the 97.5% confidence ellipse computed from the MCD estimates. We find that several observations are both one-dimensional outliers, meaning that they belong to the most extreme observations in any of the two coordinates, and correlation outliers, meaning that they do not obey the general correlation pattern between the Nasdaq and NYSE return series.

In Table 3 the sample mean and covariance matrix of the vector return series are compared with the MCD estimates. The fact that the robustly esti-

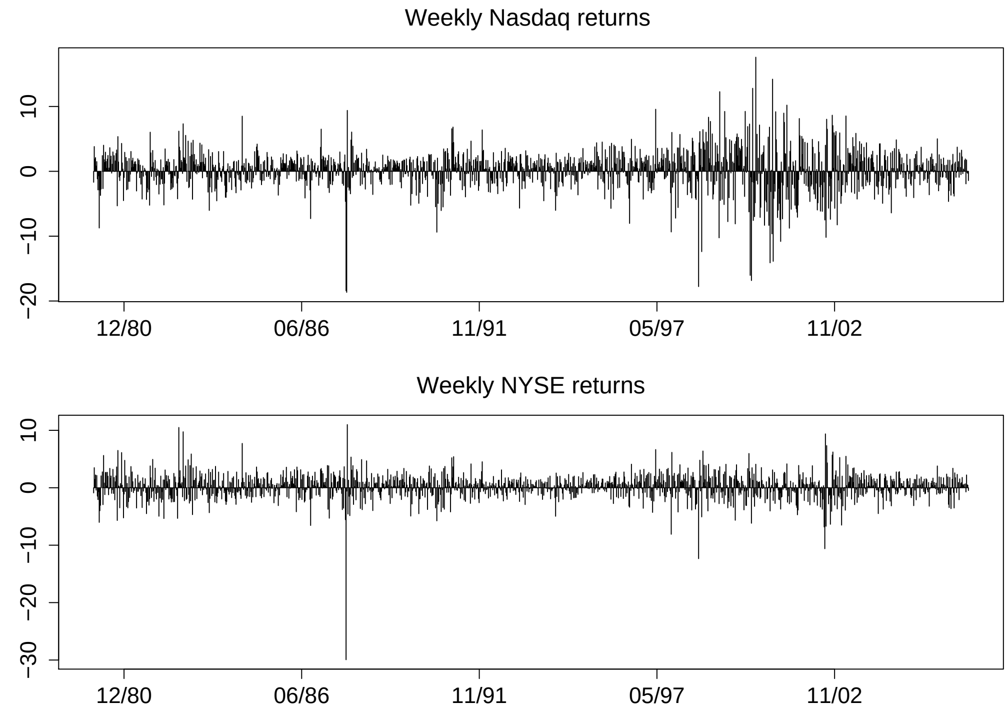

Figure 4: Weekly returns on Nasdaq and NYSE composite index.

mated mean vector is larger than the classical sample mean reflects the stylized fact in financial time series that large drawdowns in stock prices are more frequent than equally large upward movements. Note also that outliers inflate the classical sample variances and deflate the classical sample correlations (the sample correlation is 0.744 whereas the MCD correlation is 0.829).

Figure 6 reports the 0.01 trimmed lag l cross-correlations defined in Subsection 4.1 and the white noise 95% confidence bands for the components of the (absolute) return series. In the upper four plots, we see that that the return series does not contain any significant cross-correlation. Hence, the conditional location of the data can be modeled as being constant. In the third panel we see that the autocorrelation in each of the absolute return series is positive and significant. This observation can be regarded as a manifestation of volatility clustering in the two series. There seems to be a stronger persistence in the volatility of the Nasdaq return series than in the NYSE return series. The robustly estimated cross-correlation between the amplitude of the current return on one index and the magnitude of the lagged return on the other index

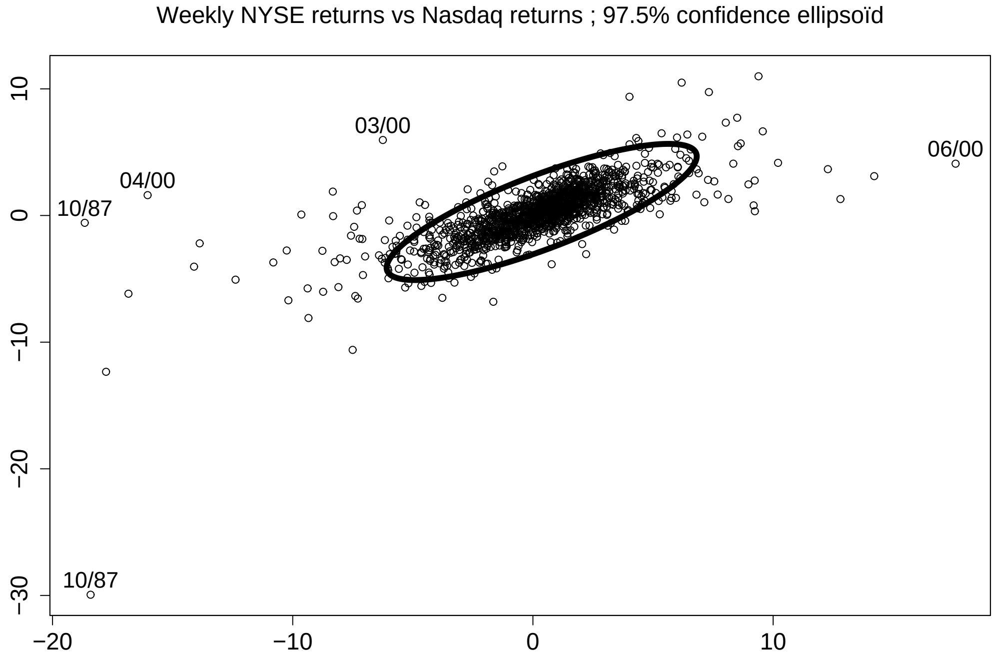

Figure 5: Scatter plot and robustly estimated 97.5% confidence ellipsoïd of weekly returns on Nasdaq and NYSE composite index. The 10 largest outliers are labeled by their time index.

is positive and significant. A multivariate volatility model is needed to decide whether this results from a combination of strong positive correlation between the magnitude of the two series and persistence in the univariate volatility process or whether it reflects genuine causality between the volatility of the two series.

Classical estimates		MCD estimates	
mean	covariance	mean	covariance
$\begin{pmatrix} 0.194 \\ 0.188 \end{pmatrix}$	$\begin{pmatrix} 9.773 & 5.283 \\ 5.283 & 5.155 \end{pmatrix}$	$\begin{pmatrix} 0.409 \\ 0.308 \end{pmatrix}$	$\begin{pmatrix} 5.906 & 4.179 \\ 4.179 & 4.287 \end{pmatrix}$

Table 3: Classical sample and robust MCD estimates for the unconditional mean and covariance matrix of the weekly returns on the Nasdaq and NYSE composite index.

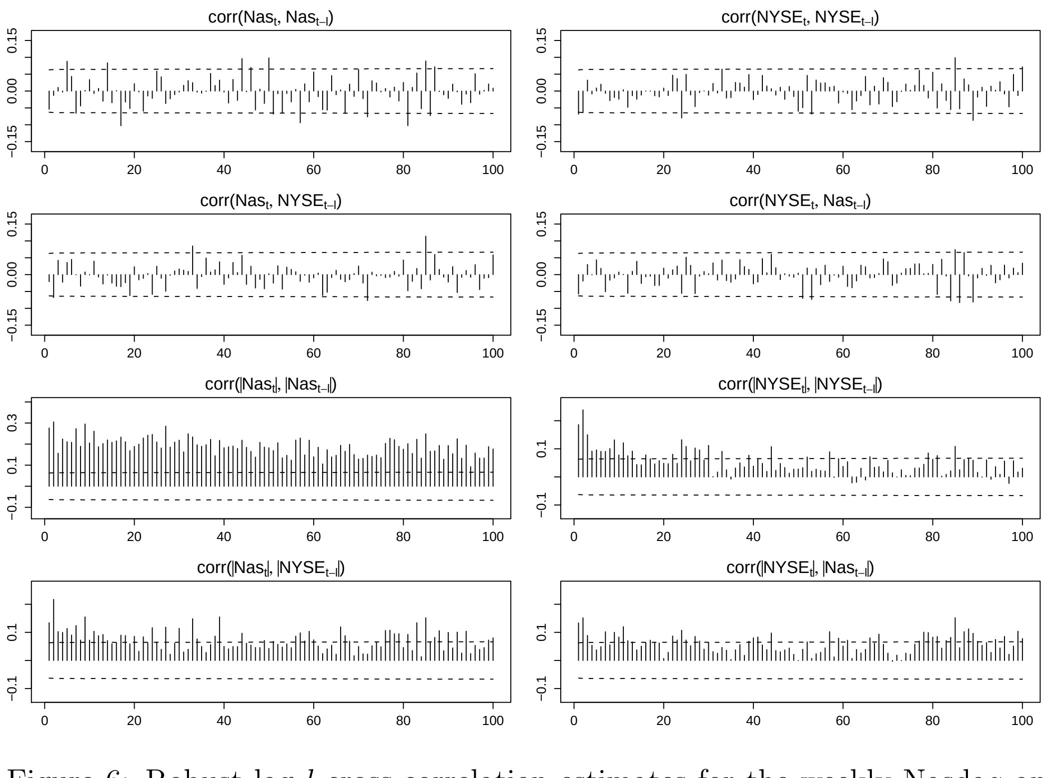

Figure 6: Robust lag l cross-correlation estimates for the weekly Nasdaq and NYSE returns (Nas$_t$, NYSE$_t$) and absolute returns ($|\text{Nas}_t|$, $|\text{NYSE}_t|$), for $l = 1, ..., 100$. The 95 % confidence bands for no cross-correlation are plotted as dashed lines.

7.2 Estimation

It is plausible to model the time series of weekly Nasdaq and NYSE returns as a conditionally Student t_ν BEKK time series process (5.1), since this model can replicate the empirical characteristics of the series, such as fat tails, absence of serial correlation in each of the return series and presence of volatility clustering. In Table 4 the P-values of the consistency factor test statistic using a M-estimator with bounded Gaussian and (un)bounded Student t_4 loss function are reported for various degrees of freedom. When the test statistic is estimated under the null that the innovations are Student t_4 distributed, the null of correct specification cannot be rejected for all considered M-estimators. For this reason, we will continue the econometric analysis of this data set assuming the returns to be conditionally Student t_4 distributed.

Loss function	ω_{ρ_ϕ}		ρ_{t_4}		$\omega_{\rho_{t_4}}$	
M / BIPM	M	BIPM	M	BIPM	M	BIPM
$H_0 : g_* = t_3$	0.00	0.00	0.75	0.46	0.00	0.00
$H_0 : g_* = t_4$	0.82	0.19	0.30	0.61	0.25	0.25
$H_0 : g_* = t_5$	0.62	0.03	0.01	0.29	0.00	0.00
$H_0 : g_* = t_\infty$	0.00	0.00	0.35	0.42	0.00	0.00

Table 4: P-values of the consistency factor test statistic for correct specification of the density function of the weekly returns on Nasdaq and NYSE indices, as estimated by the (BIP) M-estimators with bounded Gaussian loss function and (un)bounded Student t_4 loss function.

The Monte Carlo study predicts that the M-estimator with Gaussian loss function, which is better known as the Gaussian QML estimator, will be adversely affected by the outliers in the data. Table 5 reports the estimated parameters as well as their estimated asymptotic standard errors for the Gaussian QML estimator, the M-estimator with bounded Gaussian loss function and the (BIP) M-estimator with Student t_4 loss function. We see that the robust parameter estimates are similar, but very different from the estimate obtained using the non robust Gaussian QML estimator. Conform the realized volatility in the time series, the robustly estimated volatility for the Nasdaq returns is more reactive to recent variability than the estimated volatility of the NYSE return ($\hat{A}_{11} > \hat{A}_{22}$). This ranking is opposite for the Gaussian QML estimate. The robust estimates can thus be considered as more reliable.

Because of outliers and because the true density function is fat-tailed, the estimated standard errors for the Gaussian QML estimates are very large. We find that for all M-estimators, the diagonal elements of the matrices A and B are not significant. This means that the (i,j)-th element of the conditional covariance matrices $H_{t,\theta}$ and $\tilde{H}_{t,\theta}$ depends depends only on its lagged value and on $r_{i,t-1}r_{j,t-1}$. Particularly, for $i = j$, we have that news on one market has no incremental predictive power on the univariate prediction of expected volatility on the other market.

$\hat{C}_{11}$	$\hat{C}_{12}$	$\hat{C}_{22}$	$\hat{A}_{11}$	$\hat{A}_{12}$	$\hat{A}_{22}$	$\hat{B}_{11}$	$\hat{B}_{12}$	$\hat{B}_{22}$
Gaussian QML estimator								
0.645	0.546	0.270	0.334	-0.034	0.428	0.920	0.009	0.879
(0.330)	(1.085)	(2.031)	(0.797)	(0.931)	(0.168)	(0.440)	(0.529)	(0.099)
M-estimator with bounded Gaussian loss function								
0.648	0.461	0.352	0.433	-0.009	0.395	0.901	0.007	0.905
(0.413)	(0.582)	(0.094)	(0.261)	(0.253)	(0.138)	(0.169)	(0.224)	(0.042)
M-estimator with Student t_4 loss function								
0.637	0.464	0.343	0.397	-0.001	0.396	0.915	0.004	0.906
(0.079)	(0.080)	(0.152)	(0.048)	(0.075)	(0.095)	(0.017)	(0.016)	(0.034)
BIP M-estimator with Student t_4 loss function								
0.664	0.459	0.362	0.431	0.006	0.408	0.899	0.003	0.899
(0.043)	(0.038)	(0.049)	(0.034)	(0.042)	(0.046)	(0.012)	(0.011)	(0.012)

Table 5: BEKK parameter estimates for the weekly returns on the Nasdaq and NYSE composite index. Asymptotic standard errors are given in parentheses.

In Figure 7 we compare the realized returns with the 90% confidence bands for a conditionally Student t_4 BEKK and BIP-BEKK time series process as estimated by the nonrobust Gaussian QML estimator and by the robust BIP M-estimator with Student t_4 loss function, respectively. We see that there is much more time-variation in the estimated conditional volatility of the Nasdaq return series than in the NYSE return series and that it is only in certain periods of history that an investment in the Nasdaq index is much more risky than one in the NYSE index.

It is of particular interest to compare the non robustly estimated volatility and correlation (left plot) with the robust estimates (right plot) for the period of the 1987 stock market crash and the period of the burst of the internet bubble in 2000. In each of these periods, large returns of similar magnitude occur but the persistence in volatility is different. The stock market crash in 1987 was not followed by returns of similar size whereas the burst of the internet bubble corresponds to a prolonged period of high volatility. Thanks to the property of a limited effect of innovations on future volatility, the BIP-BEKK model captures well this difference in reaction to shocks. In contrast, the Gaussian QML prediction based on the BEKK model overestimates the persistence in the realized return variability in the period subsequent to the 1987 crash.

There is an extensive literature on time variation in correlation between financial assets and on the propagation of shocks across national and international markets (see e.g. Longin and Solnik, 1995). In our case, as can be seen in the bottom panel in Figure 7, the time series of the estimated conditional correlation between the Nasdaq and NYSE indices reveals that historically it is stable around the robust MCD estimate of 0.83 (and not the classical sample estimate of 0.74). This period of constant correlation has been temporarily disrupted by the large asymmetric and persistent volatility shock due to the burst of the internet bubble. Comparing the estimated conditional correlations for the time points successive to the volatility shocks, we see that, alike in the unconditional case, outliers deflate the estimated correlation.

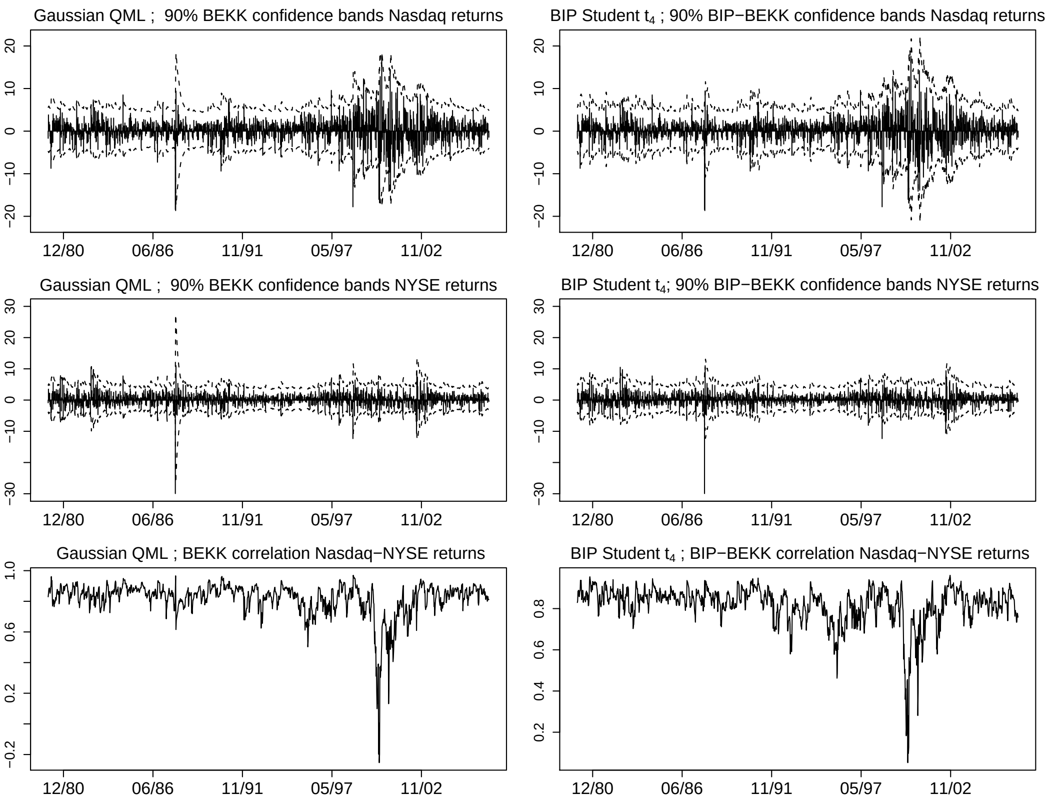

Figure 7: Weekly returns on the Nasdaq and NYSE indices and their conditional correlation. The realized returns are compared with the 90% confidence bands for a conditionally Student t_4 BEKK (left) or BIP-BEKK (right) time series process, as estimated by the Gaussian QML estimator (left) and BIP M-estimator with Student t_4 loss function (right).

8 Concluding remarks

This paper introduces a class of M-estimators for multivariate conditionally heteroscedastic time series models. We prove that under general conditions M-estimators are consistent under elliptical innovations. We study the effect of the shape of the loss function and of bounding the news impact curve on the robustness of the M-estimator to outliers. The popular Gaussian quasi-maximum likelihood estimator is shown to be very sensitive to outliers in the data. If the data is suspected to be contaminated by outliers, we recommend to use a volatility model that has the property of bounded innovation propagation and to use a M-estimator with Student t_4 loss function. The Monte Carlo study documents the good robustness properties of this approach. The definition of robust M-estimators includes a consistency factor depending on knowledge of the true density function. For this reason, we recommend to use the consistency factor test statistic, discussed in Subsection 4.2, as a validity check.

We apply this new methodology to the weekly returns on the Nasdaq and NYSE indices for the period 1987-2006. We find that the robust approach leads to more reliable parameter estimates and, in particular for the period consecutive to the outlying return, to volatility predictions that capture better the realized variability in the data. Another important empirical finding is that, in the absence of large asymmetric volatility shocks, the conditional correlation between the two indices is almost constant. We find that in contrast with the shared volatility shock in October 1987, the Nasdaq-specific volatility shock in the year 2000 has led to a persistent breakdown in the correlation for that period.

This research can be extended by looking at other classes of robust estimators and by considering other multivariate conditional heteroscedasticity models. We only studied one particular way of bounding the Gaussian and Student t loss functions and of bounding the impact of news on subsequent volatility estimates. It could be, as in Mancini *et al.* (2005), that other choices yield a better trade-off between efficiency and robustness. In the Monte Carlo study and empirical application, we focus on the bivariate symmetric BEKK volatility model. Further research is needed regarding the use and robustness of M-estimators for other types of multivariate conditional heteroscedasticity models. In multivariate GARCH models, the tendency is to reduce the dimensionality of the problem using principal component analysis or by specifying separate univariate GARCH models for the volatility of each series and a time-varying process for the conditional correlations. Further research could show how well robust methods work for these types of models.

9 Acknowledgment

The authors are grateful to Luc Bauwens, Geert Dhaene, Christian Hafner, Sébastien Laurent and Oliver Linton for very helpful comments and suggestions. This research has been supported by the Research Fund K.U.Leuven (contract number: GOA/2007/04). Scientific responsibility is assumed by the authors.

A Consistency of the M-estimator

Denote $m(r_t;\theta,\rho) = \log\det H_{t,\theta} + \sigma_*\rho\left(r_t' H_{t,\theta}^{-1} r_t\right)$. Its first partial derivative $\dot{m}_i(r_t;\theta,\rho) = \partial m(r_t;\theta,\rho)/\partial\theta_i$ is given by

$$\operatorname{Tr}\left[I_N - \sigma_*\psi(r_t' H_{t,\theta}^{-1} r_t) r_t r_t' H_{t,\theta}^{-1}\right] \frac{\partial H_{t,\theta}}{\partial\theta_i} H_{t,\theta}^{-1}, \tag{A.1}$$

(Comte and Lieberman, 2003). The square N-dimensional matrix $\partial H_{t,\theta}/\partial\theta_i$ holds the i-th partial derivative of the corresponding elements of $H_{t,\theta}$ and Tr is the trace operator. Since $\dot{m}_i(r_t;\theta,\rho)$ is a measurable function of the strictly stationary and ergodic process $\{r_t\}$, it is also strictly stationary and ergodic. By the Ergodic Theorem its time series average, which is the score of the M-function, will converge to its ensemble mean. We thus find that the asymptotic limit of the score of the M-function equals

$$\operatorname{Tr}\ E_{\theta_*}\left\{\left[I_N - \sigma_*\psi(r_t' H_{t,\theta}^{-1} r_t) r_t r_t' H_{t,\theta}^{-1}\right] \frac{\partial H_{t,\theta}}{\partial\theta_i} H_{t,\theta}^{-1}\right\}. \tag{A.2}$$

Under the no dominance assumption A5, convergence in probability will be uniform.

We will now show that under A1-A6, the limit of the M-function has a unique extremum in θ_*. Indeed, Maronna (1976) proves that under A1, there exists a unique solution to the following implicit equation determining X:

$$E_{\theta_*}\left[\sigma_*\psi(r_t' X^{-1} r_t) r_t r_t' | I_{r_{t-1}}\right] = X. \tag{A.3}$$

The solution to this equation is called the M-estimator of scatter of r_t or also the pseudo-covariance matrix of r_t. Since this estimator is affine equivariant and since r_t is assumed to be elliptically symmetric with conditional scatter matrix H_{t,θ_*}, one can show that X will be proportional to the true conditional scatter matrix H_{t,θ_*}, with proportionality factor c equal to

$$c = \frac{N}{\sigma_*}\left\{\frac{2\pi^{N/2}}{\Gamma(N/2)}\int_0^\infty \psi(t^2) t^{N+1} g_*(t^2) dt\right\}^{-1}, \tag{A.4}$$

(Chapter 13 in Bilodeau and Brenner, 1999, Chapter 6 in Maronna *et al.*, 2006). It follows that M-estimators will be consistent since σ_* in (2.8) is such that $c = 1$. By the no observational equivalence assumption A3, it is only for θ_* that the conditional covariance matrix equals H_{t,θ_*}. We may thus conclude that for loss functions satisfying A1, the limit of the M-function will have an extremum in θ_*. This extremum is unique under the assumption A6 that there is no $\theta \in \Theta$ for which the score of the M-function is zero while the conditional expectation under $\mathcal{P}_{\theta_*}$ of the weighted covariance matrix is not the true conditional covariance matrix H_{t,θ_*}. A further sufficient condition to ensure that the extremum of the limit is the limit of the extremum of the M-estimator is that the parameter space Θ is compact.

References

[1] Aslihan AS, Pinar MC, Leyffer S. 2003. Constrained nonlinear programming for volatility estimation with GARCH models. *SIAM review* **45**: 485-503.

[2] Bauwens L, Laurent S, Rombouts JVK. 2006. Multivariate GARCH models: a survey. *Journal of Applied Econometrics* **21**: 79-109.

[3] Bilodeau M, Brenner D. 1999. *Theory of Multivariate Statistics*. Springer, New York.

[4] Chan WS, Wei WWS. 1992. A comparison of some estimators of time series autocorrelation. *Computational Statistics and Data Analysis* **14**: 149-163.

[5] Comte F, Lieberman O. 2003. Asymptotic theory for multivariate GARCH processes. *Journal of Multivariate Analysis* **84**: 61-84.

[6] Engle RF, Kroner FK. 1995. Multivariate simultaneous generalized ARCH. *Econometric Theory* **11**: 122-150.

[7] Engle RF, Ng VK. 1993. Measuring and testing the impact of news on volatility. *The Journal of Finance* **58**: 1749-1778.

[8] Fiorentini G, Sentana E, Calzolari G. 2003. Maximum likelihood estimation and inference in multivariate conditionally heteroscedastic dynamic regression models with Student t innovations. *Journal of Business and Economic Statistics* **21**: 532-546.

[9] Franses P, Ghijsels H. 1999. Additive outliers, GARCH and forecasting volatility. *International Journal of Forecasting* **15**: 1-9.

[10] Hafner CM, Herwartz H. 2003. Analytical quasi maximum likelihood inference in multivariate volatility models. Econometric Institute Report 21. Erasmus University Rotterdam.

[11] Hayashi F. 2000. *Econometrics*. Princeton University Press: Princeton.

[12] Jeantheau T. 1998. Strong consistency of estimators for multivariate ARCH models. *Econometric Theory* **14**: 70-86.

[13] Longin F, Solnik B. 1995. Is the correlation in international equity returns constant: 1960-1990? *Journal of International Money and Finance* **4**: 3-26.

[14] Mancini L, Ronchetti E, Trojani F. 2005. Optimal conditionally unbiased bounded-influence inference in dynamic location and scale models. *Journal of the American Statistical Association* **105**: 628-641.

[15] Maronna RA. 1976. Robust M-estimators of multivariate location and scatter, *The Annals of Statistics* **4**: 51-67.

[16] Maronna RA, Martin RD, Yohai VJ. 2006. *Robust Statistics.* Wiley, New York.

[17] Muler N, Yohai VJ. 2002. Robust estimates for ARCH processes, *Journal of Time Series Analysis* **23**: 79-109.

[18] Muler N, Yohai VJ. 2006. Robust estimates for GARCH models. Preprints Instituto de Calculo, University of Buenos Aires.

[19] Newey WK, Steigerwald DG. 1997. Asymptotic bias for quasi-maximum-likelihood estimators in conditional heteroskedasticity models. *Econometrica* **65**: 587-599.

[20] Nocedal J, Wright SJ. 2000. *Numerical Optimization.* Springer Series in Operations Research. Springer Verlag, New York.

[21] Park B. 2002. An outlier robust GARCH model and forecasting volatility of exchange rate returns. *Journal of Forecasting* **21**: 381-393.

[22] Peng L, Yao Q. 2003. Miscellanea. Least absolute deviations estimation for ARCH and GARCH models. *Biometrika* **90**: 967-997.

[23] Rousseeuw PJ. 1985. Multivariate estimation with high breakdown point. In *Mathematical Statistics and Its Applications (vol. B)*, Grossmann W, Pflug G, Vincze I and Wertz W (eds). Dordrecht-Reidel.

[24] Rousseeuw PJ, van Zomeren BC. 1990. Unmasking multivariate outliers and leverage points. *Journal of the American Statistical Association* **85**: 633-639.

[25] Sakata S, White H. 1998. High breakdown point conditional dispersion estimation with application to S&P 500 daily returns volatility. *Econometrica* **66**: 529-567.

[26] Tiao GC, Box GEP. 1981. Modeling multiple time series with applications. *Journal of the American Statistical Association* **76**: 802-816.

[27] Van Dijk D, Franses PH, Lucas A. 1999. Testing for ARCH in the presence of additive outliers. *Journal of Applied Econometrics* **14**: 539-562.